

Biostatistics (2012), **13**, 3, pp. 371–383

doi:10.1093/biostatistics/kxr032

Advance Access publication on October 31, 2011

Competing risks regression for clustered data

BINGQING ZHOU*

*Division of Biostatistics, School of Public Health, Yale University, New Haven, CT 06520, USA
and VA Cooperative Studies Program Coordinating Center, West Haven, CT 06516, USA*

bingqing.zhou@yale.edu

JASON FINE

*Department of Biostatistics and Department of Statistics, University of North Carolina,
Chapel Hill, NC 27599, USA*

AURELIEN LATOUCHE

Conservatoire national des arts et metiers and INSERM U1018, 75003 Paris, France

MYRIAM LABOPIN

*European Blood and Marrow Transplant Group, Hopital Saint Antoine, Paris University,
75010 Paris, France*

SUMMARY

A population average regression model is proposed to assess the marginal effects of covariates on the cumulative incidence function when there is dependence across individuals within a cluster in the competing risks setting. This method extends the Fine–Gray proportional hazards model for the subdistribution to situations, where individuals within a cluster may be correlated due to unobserved shared factors. Estimators of the regression parameters in the marginal model are developed under an independence working assumption where the correlation across individuals within a cluster is completely unspecified. The estimators are consistent and asymptotically normal, and variance estimation may be achieved without specifying the form of the dependence across individuals. A simulation study evidences that the inferential procedures perform well with realistic sample sizes. The practical utility of the methods is illustrated with data from the European Bone Marrow Transplant Registry.

Keywords: Clustered; Competing risk; Hazard of subdistribution; Marginal model; Martingale; Multivariate; Partial likelihood.

1. INTRODUCTION

Competing risks regression was proposed by Fine and Gray (1999) to assess the effects of covariates on the probabilities of a particular cause of failure with independent subjects, where each subject may fail from one of several causes. Naively, censoring subjects who fail from causes other than that of interest may yield an invalid analysis for the cumulative probabilities of the cause of interest across time. The Fine and Gray (1999) regression model, which is specified for the cumulative incidence, treats the competing

*To whom correspondence should be addressed.

causes differently from independent censoring variables and has been adopted in practice as an alternative to the naive analysis described above.

In many applications involving competing risks, individuals may be correlated within clusters, owing to unobserved shared factors across individuals. We refer to such data as “clustered competing risks,” with “clustering” referring to the potential dependence across individuals within clusters and “competing risks” referring to the potential dependence across causes within individuals, analogously to the case of independent subjects (Fine and Gray, 1999). In family studies of hereditary cancer, parents and children may share genetic and environmental factors, leading to familial correlations in disease onset. In multicenter studies, patient population and referral pattern in each center may result in correlated outcomes within centers. The current work extends the Fine and Gray (1999) model to this clustered setup.

Numerous semiparametric regression models have been proposed for the clustered data setting with independent censoring but without dependent competing risks. Wei *and others* (1989) proposed a proportional hazards model with common covariate effects but different baseline hazard for each cluster member. Lee *and others* (1992), Liang *and others* (1993), and Cai and Prentice (1997) studied proportional hazards models with common baseline and regression coefficients. These marginal models are specified unconditionally, with both the regression coefficients and the baseline hazards having population average interpretations. Inferences for the marginal regression models are generally robust to assumptions about the within-cluster correlations. To investigate the covariate effects conditionally on cluster, frailty models may be employed (Clayton and Cuzick 1985; Hougaard 1986). These conditional models involve explicit assumptions on the within-cluster dependence, enabling an assessment of both covariate effects and within-cluster associations.

For clustered competing risks, one may employ methods for independently censored data described above in the analysis of the cause-specific hazard functions (Prentice *and others*, 1978). These methods are not generally appropriate for the cumulative incidence function. Existing methods for the cumulative incidence with clustered data include a random effects Fine–Gray model specified conditionally on cluster (Katsahian *and others*, 2006) and modifications of the widely used nonparametric Gray’s (1988) test (Chen *and others*, 2008), which is applicable with categorical covariates. If the main goal is formulating a cumulative incidence regression model to investigate the effects of covariates, introducing a dependence parameter to model associations within individuals in a cluster does not seem to have much advantage over a marginal modeling strategy, in which the dependence is unspecified.

In this article, we propose to extend the Fine–Gray model to the clustered data setting, so that the cumulative incidence function can be estimated by adjusting for prognostic factors while accommodating correlation within clusters. We construct a marginal proportional subdistribution hazards model, which is similar to the Lee *and others* (1992) marginal model except that we focus on subdistribution hazards instead of cause-specific hazards. Under an independence working assumption, the cumulative incidence function and the effects of the prognostic factors can be estimated by following the Fine–Gray methodology. The proposed variance estimator accommodates the correlation within clusters. *Ad hoc* approaches to such marginal competing risks models might be utilized, with Ruan and Gray (2008) and Logan *and others* (2011) employing imputation and pseudovalues at discrete time points, respectively. However, the current study is the first to formally investigate inferential issues.

Section 2 introduces the marginal Fine–Gray model for clustered data. In Section 3, we discuss partial likelihood inferences for the marginal model, as well as inverse weighted estimating equations, which permit independent censoring from loss to follow-up in addition to the potentially dependent competing risks. Such inverse weighting methods are nonstandard, as they involve estimating the censoring probabilities using clustered censoring times, which may be correlated. Asymptotic properties of the estimators are established, along with a simple plug-in variance estimator. Section 4 assesses the performance of the proposed model through simulation studies, with the analysis of the motivating data following in

Section 5. We conclude with a few remarks. The technical details are included in the supplementary material available at *Biostatistics* online.

2. MARGINAL PROPORTIONAL SUBDISTRIBUTION HAZARDS MODEL

Let T_{ik} be the failure time for the k th member in the i th cluster, $i = 1, \dots, n; k = 1, \dots, m_i$. Let $\varepsilon_{ik} \in \{1, \dots, l\}$ denote the corresponding cause of failure and $\mathbf{Z}_{ik} = \{Z_{1ik}, \dots, Z_{pik}\}$ be a $p \times 1$ vector of covariates. For right-censored data, one observes $\{X_{ik} = T_{ik} \wedge C_{ik}, \xi_{ik} = I(T_{ik} \leq C_{ik})\varepsilon_{ik}, \mathbf{Z}_{ik}\}$, where C_{ik} is the independent censoring time, $\xi_{ik} \in \{0, 1, \dots, l\}$, and $a \wedge b = \min(a, b)$. The covariates \mathbf{Z}_{ik} may include time-varying components, which are known deterministic functions of time and/or time-independent components, hence fully observed. The dependence on time is suppressed where possible, to ease readability.

We assume that $(T_{ik}, \varepsilon_{ik})$ and C_{ik} are independent given \mathbf{Z}_{ik} for each i and k . Let $\mathbf{T}_i = (T_{i1}, \dots, T_{im_i})$, $\varepsilon_i = (\varepsilon_{i1}, \dots, \varepsilon_{im_i})$, $\mathbf{Z}_i = (\mathbf{Z}_{i1}, \dots, \mathbf{Z}_{im_i})$, and $\mathbf{C}_i = (C_{i1}, \dots, C_{im_i})$. We assume that $(\mathbf{T}_i, \varepsilon_i, \mathbf{Z}_i, \mathbf{C}_i, m_i)_{i=1, \dots, n}$ are i.i.d. We also require that $(\mathbf{T}_i, \varepsilon_i)$ and \mathbf{C}_i are independent given (\mathbf{Z}_i, m_i) . In cluster i , the components of $(\mathbf{T}_i, \varepsilon_i)$ may be correlated conditionally on (\mathbf{Z}_i, m_i) , and similarly, the components of \mathbf{C}_i may be dependent conditionally on (\mathbf{Z}_i, m_i) . Let $\mathbf{X}_i = (X_{i1}, \dots, X_{im_i})$ and $\xi_i = (\xi_{i1}, \dots, \xi_{im_i})$. The observed data within each cluster, $(\mathbf{X}_i, \xi_i, \mathbf{Z}_i, m_i)$, are assumed to be i.i.d. across clusters $i = 1, \dots, n$.

We are interested in assessing the effects of covariates on the marginal cumulative incidence function for failure from cause 1, conditional on the covariates: $F_1(t; \mathbf{Z}_{ik}) \equiv \Pr(T_{ik} \leq t, \varepsilon_{ik} = 1 | \mathbf{Z}_{ik})$. The marginal subdistribution hazard $\lambda_1(t; \mathbf{Z}_{ik}) = dF_1(t; \mathbf{Z}_{ik}) / \{1 - F_1(t; \mathbf{Z}_{ik})\}$ is modeled as

$$\lambda_1(t; \mathbf{Z}_{ik}) = \lambda_{10}(t) \exp(\beta'_0 \mathbf{Z}_{ik}), \quad (2.1)$$

where $\lambda_{10}(\cdot)$ is a completely unspecified baseline subdistribution hazard function and β_0 is a $p \times 1$ vector of unknown regression parameters. As in Fine and Gray (1999), our methodology does not require any assumptions regarding the cumulative incidence functions for other causes. As noted in Fine and Gray (1999), there is potential to gain efficiency by adding such assumptions but with greatly increased computational and theoretical challenges. Moreover, our analyses of model (2.1) will not be affected by whether we treat all competing causes $\{2, \dots, l\}$ separately or group them. It is worth emphasizing that independent censoring by C enters differently into the analysis than the competing events and cannot be grouped with competing causes.

3. ESTIMATION AND INFERENCE

3.1 Censoring complete data

The term censoring complete (CC) data refers to the case in which failure time T is right censored but potential censoring time C is always observed, that is, censoring results only from administrative loss to follow-up. We modify the partial likelihood of Fine and Gray (1999) to accommodate clustering via an independence working assumption. The case of complete data where C is larger than the maximal failure time, that is, all event times and types are observed, is a special case of this setup. Hereafter, it is assumed that there exists a τ such that $P(T_{ik} > \tau) > \delta > 0$, $P(C_{ik} = \tau) = P(C_{ik} \geq \tau) > \delta > 0$ for all i, k .

Let $N_{ik}(t) = I(T_{ik} \leq t, \xi_{ik} = 1)$ and $Y_{ik}(t) = 1 - N_{ik}(t-)$ denote the counting process and risk process for the complete data, respectively. When the data are CC, the risk process is modified to $Y_{ik}^*(t) = I(C_{ik} \geq t)Y_{ik}(t)$.

For convenience, we define the following notation:

$$\mathbf{S}^{(r)}(\beta, t) = n^{-1} \sum_{i=1}^n \sum_{k=1}^{m_i} I(C_{ik} \geq t) Y_{ik}(t) \mathbf{Z}_{ik}(t)^{\otimes r} e^{\beta' \mathbf{Z}_{ik}(t)}, \quad r = 0, 1, 2,$$

$$\mathbf{s}^{(r)}(\beta, t) = \lim_{n \rightarrow \infty} \mathbf{S}^{(r)}(\beta, t),$$

$$\bar{\mathbf{Z}}(\beta, t) = \mathbf{S}^{(1)}(\beta, t) / S^{(0)}(\beta, t), \quad \mathbf{V}(\beta, t) = \frac{\mathbf{S}^{(2)}(\beta, t)}{S^{(0)}(\beta, t)} - \frac{\mathbf{S}^{(1)}(\beta, t)^{\otimes 2}}{S^{(0)}(\beta, t)^2},$$

$$\mathbf{e}(\beta, t) = \mathbf{s}^{(1)}(\beta, t) / s^{(0)}(\beta, t), \quad \mathbf{v}(\beta, t) = \frac{\mathbf{s}^{(2)}(\beta, t)}{s^{(0)}(\beta, t)} - \frac{\mathbf{s}^{(1)}(\beta, t)^{\otimes 2}}{s^{(0)}(\beta, t)^2},$$

where $\mathbf{a}^{\otimes 0} = 1$, $\mathbf{a}^{\otimes 1} = \mathbf{a}$, and $\mathbf{a}^{\otimes 2} = \mathbf{a}\mathbf{a}^T$. The following regularity conditions are needed:

ASSUMPTION 3.1 Regularity conditions assumed throughout:

- A1. $\int_0^\tau \lambda_{10}(t) dt < \infty$.
- A2. $Z_{pik}(\cdot)$ have bounded total variations, that is, $|Z_{pik}(0)| + \int_0^\tau |dZ_{pik}(t)| \leq M$ for all p, i , and k , where M is a constant.
- A3. There exists a neighborhood \mathcal{B} of β_0 and scalar, vector, and matrix functions $s^{(0)}$, $\mathbf{s}^{(1)}$, and $\mathbf{s}^{(2)}$, defined on $\mathcal{B} \times [0, \tau]$ such that for $r = 0, 1, 2$, $\sup_{t \in [0, \tau], \beta \in \mathcal{B}} \|\mathbf{S}^{(r)}(\beta, t) - \mathbf{s}^{(r)}(\beta, t)\|$ converges in probability to zero.
- A4. $\mathbf{s}^{(r)}(\beta, t)$ are continuous functions of $\beta \in \mathcal{B}$ uniformly in $t \in [0, \tau]$ and are bounded on $\mathcal{B} \times [0, \tau]$. $\mathbf{s}^{(r)}(\beta, t)$ are bounded away from zero.
- A5. The matrix $\Omega = \int_0^\tau \mathbf{v}(\beta_0, u) s^{(0)}(\beta_0, u) \lambda_{10}(u) du$ is positive definite.

Under the independence working assumption, the pseudopartial likelihood (Lee *and others*, 1992) for β_0 is

$$L^*(\beta) = \prod_{i=1}^n \prod_{k=1}^{m_i} \left\{ \frac{e^{\beta' \mathbf{Z}_{ik}(X_{ik})}}{\sum_{i'=1}^n \sum_{k'=1}^{m_{i'}} I(C_{i'k'} \geq X_{ik}) Y_{i'k'}(X_{ik}) e^{\beta' \mathbf{Z}_{i'k'}(X_{ik})}} \right\}^{I(\xi_{ik}=1)},$$

with corresponding estimating function

$$\mathbf{U}_1^*(\beta) = \sum_{i=1}^n \sum_{k=1}^{m_i} \int_0^\tau \{\mathbf{Z}_{ik}(u) - \bar{\mathbf{Z}}(\beta, u)\} I(C_{ik} \geq u) dN_{ik}(u). \quad (3.1)$$

The estimator $\hat{\beta}$ that maximizes $L^*(\beta)$ may be obtained as a solution to $\mathbf{U}_1^*(\beta) = 0$. The asymptotic results for CC data estimation are natural extensions of those from the ordinary Cox-type model for clustered data (Lee *and others* 1992; Spiekerman and Lin 1998). We present the main results briefly, as a point of comparison for subsequent results where inverse weighting is needed. In the supplementary

material available at *Biostatistics* online, the estimator $\hat{\beta}$ is shown to be consistent and asymptotically normal, with variance which may be consistently estimated using a sandwich variance estimator, which is robust to the within-cluster correlations. The form of the variance estimator is essentially identical to that in Lee and others (1992), with a minor modification for the censoring time always being observed.

3.2 Right-censored data

When the data are right censored, inverse probability of censoring weighting (Robins and Rotnitzky, 1992) techniques cannot be applied directly, owing to correlation within clusters. To account for such clustering, we define a marginal inverse probability of censoring weight for subject k in the i th cluster at time t . As in Fine and Gray (1999), we assume that C_{ik} is independent of \mathbf{Z}_{ik} , with the weight $w_{ik}(t) \equiv I(C_{ik} \geq T_{ik} \wedge t)G(t)/G(X_{ik} \wedge t)$, where $G(t) = \Pr(C_{ik} \geq t)$.

Since the cluster sizes are finite, we cannot consistently estimate the censoring distribution in each cluster. Hence, we pool across cluster via the assumption of a single G . Note that the independent censoring times may be correlated within a cluster, but that data are uncorrelated across clusters.

We naively estimate $G(\cdot)$ with $\hat{G}(\cdot)$, the Kaplan–Meier estimator of the survival function of the censoring random variable in which the dependence among individuals within clusters is ignored. The estimated weight $\hat{w}_{ik}(t) = I(C_{ik} \geq T_{ik} \wedge t)\hat{G}(t)/\hat{G}(X_{ik} \wedge t)$ is used in place of $w_{ik}(t)$ in the estimating equation. In the case where $m_i = 1$; $i = 1, \dots, n$, this approach is equivalent to that in Fine and Gray (1999).

In addition to the notation in Section 3.1, the following notation is needed.

$$\hat{\mathbf{S}}^{(r)}(\beta, t) = n^{-1} \sum_{i=1}^n \sum_{k=1}^{m_i} \hat{w}_{ik}(t) Y_{ik}(t) \mathbf{Z}_{ik}(t)^{\otimes r} e^{\beta' \mathbf{Z}_{ik}(t)}, \quad r = 0, 1, 2,$$

$$\hat{\mathbf{Z}}(\beta, t) = \frac{\hat{\mathbf{S}}^{(1)}(\beta, t)}{\hat{\mathbf{S}}^{(0)}(\beta, t)}, \quad \hat{\mathbf{V}}(\beta, t) = \frac{\hat{\mathbf{S}}^{(2)}(\beta, t)}{\hat{\mathbf{S}}^{(0)}(\beta, t)} - \frac{\hat{\mathbf{S}}^{(1)}(\beta, t)^{\otimes 2}}{\hat{\mathbf{S}}^{(0)}(\beta, t)^2},$$

$$S^c(t) = n^{-1} \sum_{i=1}^n \sum_{k=1}^{m_i} I(X_{ik} \geq t), \quad \pi(t) = \lim_{n \rightarrow \infty} S^c(t).$$

The inverse weighted estimating equation is

$$\mathbf{U}_1(\beta) = \sum_{i=1}^n \sum_{k=1}^{m_i} \int_0^\tau \{\mathbf{Z}_{ik}(u) - \hat{\mathbf{Z}}(\beta, u)\} \hat{w}_{ik}(u) dM_{ik}(\beta, u), \quad (3.2)$$

where $M_{ik}(\beta, t) = \int_0^t dN_{ik}(u) - \int_0^t Y_{ik}(u) \lambda_{10}(u) e^{\beta' \mathbf{Z}_{ik}(u)} du$ is a martingale for marginal complete data filtration $\mathcal{F}_{ik}(t) = \sigma\{N_{ik}(u), Y_{ik}(u), Y_{ik}(u) \mathbf{Z}_{ik}(u), u \leq t\}$ for each i, k but not a martingale for the joint filtration due to intracluster correlation.

The estimator $\hat{\beta}$ is obtained by solving $\mathbf{U}_1(\beta) = 0$. The consistency and asymptotic normality of $\hat{\beta}$ and a consistent variance estimator are given in a sequence of lemmas and theorems in Supplementary Appendix B of the supplementary material available at *Biostatistics* online.

A key step in the proof is deriving the asymptotic properties of the naive Kaplan–Meier estimator \hat{G} , accounting for correlations within clusters. Ying and Wei (1994) showed that the Kaplan–Meier estimator for dependent and possibly censored failure times is consistent. However, in their work, the corresponding censoring times are independent and nonrandom. In our situation, the censoring times for the censoring time random variable, that is, the failure times, are correlated within cluster. Thus, their results do not apply

directly here. In Supplementary Appendix B.1. of the supplementary material available at *Biostatistics* online, we showed $\hat{G}(\cdot)$ converges in probability to $G(\cdot)$ uniformly on $[0, \tau]$ and $n^{1/2}\{\hat{G} - G\}$ converges weakly to a tight Gaussian process with covariance function $\Sigma^c(s, t) = E\{I_i^c(s)I_i^c(t)\}$, where $I_i^c(t) = \sum_{k=1}^{m_i} \int_0^t \{\pi(u)\}^{-1} M_{ik}^c(u)$, $M_{ik}^c(t) = N_{ik}^c(t) - \int_0^t I(X_{ik} \geq u) d\Lambda_0^c(u)$ is a martingale for the marginal complete data censoring filtration.

The variance of the limiting normal distribution for $n^{1/2}(\hat{\beta} - \beta_0)$ is $\Omega^{-1} \Sigma \Omega^{-1}$, where Ω , defined in Assumption 3.1, is the limit of the negative of the partial derivative matrix of $n^{-1}U_1(\beta)$ evaluated at β_0 ; and $\Sigma = E\{(\eta_1 + \psi_1)^{\otimes 2}\}$ is the variance of asymptotic normal distribution of $n^{-1/2}U_1(\beta_0)$, $\eta_{ik} = \int_0^\tau \{\mathbf{Z}_{ik}(u) - \mathbf{e}(\beta_0, u)\} w_{ik}(u) dM_{ik}(\beta_0, u)$, $\mathbf{q}(u) = -\lim_{n \rightarrow \infty} n^{-1} \sum_{i=1}^n \sum_{k=1}^{m_i} \int_0^\tau \{\mathbf{Z}_{ik}(t) - \mathbf{e}(\beta_0, t)\} w_{ik}(t) I(X_{ik} < u \leq t) dM_{ik}(t, \beta_0)$, and $\psi_{ik} = \int_0^\infty \mathbf{q}(u) / \pi(u) dM_{ik}^c(u)$.

A consistent estimator may be obtained with $\hat{\Omega}^{-1} \hat{\Sigma} \hat{\Omega}^{-1}$, where

$$\hat{\Omega} = \sum_{i=1}^n \sum_{k=1}^{m_i} I(\xi_{ik} = 1) \left\{ \frac{\hat{\mathbf{S}}^{(2)}(\hat{\beta}, X_{ik})}{\hat{\mathbf{S}}^{(0)}(\hat{\beta}, X_{ik})} - \frac{\hat{\mathbf{S}}^{(1)}(\hat{\beta}, X_{ik})^{\otimes 2}}{\hat{\mathbf{S}}^{(0)}(\hat{\beta}, X_{ik})^2} \right\},$$

$$\hat{\Sigma} = n^{-1} \sum_{i=1}^n (\hat{\eta}_i + \hat{\psi}_i)^{\otimes 2},$$

$$\hat{\eta}_{ik} = I(\xi_{ik} = 1) \{\mathbf{Z}_{ik}(X_{ik}) - \hat{\mathbf{Z}}(\hat{\beta}, X_{ik})\} - \sum_{i'=1}^n \sum_{k'=1}^{m_i} I(\xi_{i'k'} = 1)$$

$$\times \{\mathbf{Z}_{ik}(X_{i'k'}) - \hat{\mathbf{Z}}(\hat{\beta}, X_{i'k'})\} \hat{w}_{ik}(X_{i'k'}) Y_{ik}(X_{i'k'}) e^{\hat{\beta}' \mathbf{Z}_{ik}(X_{i'k'})} \{n \hat{\mathbf{S}}^{(0)}(\hat{\beta}, X_{i'k'})\}^{-1}.$$

$$\hat{\psi}_{ik} = I(\xi_{ik} = 0) \frac{\hat{\mathbf{q}}(X_{ik})}{S^c(X_{ik})} - \sum_{i'=1}^n \sum_{k'=1}^{m_i} I(\xi_{i'k'} = 0) I(X_{ik} \geq X_{i'k'}) \frac{\hat{\mathbf{q}}(X_{i'k'})}{S^c(X_{i'k'})} \frac{1}{n S^c(X_{i'k'})}, \text{ and}$$

$$\hat{\mathbf{q}}(u) = n^{-2} \sum_{i'=1}^n \sum_{k'=1}^{m_i} \sum_{i=1}^n \sum_{k=1}^{m_i} \{\mathbf{Z}_{ik}(X_{i'k'}) - \hat{\mathbf{Z}}(\hat{\beta}, X_{i'k'})\} I(\xi_{i'k'} = 1) I(\xi_{ik} > 1)$$

$$\times I(X_{ik} < u \leq X_{i'k'}) \frac{\hat{G}(X_{i'k'})}{\hat{G}(X_{ik})} e^{\hat{\beta}' \mathbf{Z}_{ik}(X_{i'k'})} / \hat{\mathbf{S}}^{(0)}(\hat{\beta}, X_{i'k'}).$$

The \cdot in the subscripts indicates summing over all the subjects represented by that index. These results can be used to construct confidence intervals and to conduct hypothesis tests about β_0 .

To predict the cumulative incidence at a time t for a patient with covariates \mathbf{Z}_0 for right-censored data, we estimate the cumulative hazard $\Lambda_1(t, \mathbf{Z}_0) = \int \lambda_1(t, \mathbf{Z}_0) dt$ by

$$\begin{aligned} \hat{\Lambda}_1(t, \mathbf{Z}_0) &= \int_0^t \exp\{\hat{\beta}' \mathbf{Z}_0\} d\hat{\Lambda}_{10}(u) \\ &= \frac{1}{n} \sum_{i=1}^n \sum_{k=1}^{m_i} \int_0^t \frac{\exp\{\hat{\beta}' \mathbf{Z}_0\}}{\hat{\mathbf{S}}^{(0)}(\hat{\beta}, u)} \hat{w}_{ik}(u) dN_{ik}(u). \end{aligned}$$

where $\hat{\beta}$ is the estimator obtained previously in this subsection. The conditional cumulative incidence estimator $\hat{F}_1(t, \mathbf{Z}_0) = 1 - \exp\{-\hat{\Lambda}_1(t, \mathbf{Z}_0)\}$.

One can establish the consistency and asymptotic normality of $\hat{F}_1(t, \mathbf{Z}_0)$ using methods similar to those in Fine and Gray (1999). The variance is rather complicated, with bootstrapping providing practicable inferences for $F_1(t, \mathbf{Z}_0)$.

4. SIMULATION STUDIES

4.1 Data generation

Numerical investigations were conducted to assess the performance of the proposed weighted estimation approach. In the following, there are 2 causes of failure. Let the subdistribution hazard for cause 1 conditionally on a frailty v_i for cluster i satisfy

$$\mu_1(t; \mathbf{Z}_{ik}, v_i) = v_i \mu_{10}(t) \exp(\tau_1' \mathbf{Z}_{ik}). \quad (4.1)$$

To ensure that the cumulative incidence model unconditionally on v_i also satisfies proportional subdistribution hazard, we take v_i 's to be a random sample from a positive stable distribution with parameter α . The relationship between models (2.1) and (4.1) is: $\lambda_{10}(t) = \alpha M_0^{(\alpha-1)}(t) \mu_{10}(t)$, where $M_0(t) = \int_0^t \mu_{10}(u) du$ and $\beta_1 = \alpha \tau_1$. In the sequel, we let $\mu_{10}(t) = \rho e^{-\rho t}$, $\rho > 0$ such that $M_0(\infty) = 1$.

Let $F_1(t; \mathbf{Z}_{ik}, v_i) \equiv P(T_{ik} \leq t, \varepsilon_{ik} = 1 | \mathbf{Z}_{ik}, v_i)$. By the Laplace transformation (Hougaard, 1986; Logan and others, 2011),

$$F_1(t; \mathbf{Z}_{ik}) = \int_0^t F_1(u; \mathbf{Z}_{ik}, v_i) dF_{V_i}(u) = 1 - \exp\{-M_0^\alpha(t) e^{\beta_1' \mathbf{Z}_{ik}}\}.$$

An exponential distribution for $P(T_{ik} | \varepsilon_{ik} = 2, \mathbf{Z}_{ik}, h_i)$ is assumed with subdistribution hazard function $h_i \exp(\tau_2' \mathbf{Z}_{ik})$, where h_i is generated from a positive stable distribution (γ) for each i . Thus, $P(T_{ik} | \varepsilon_{ik} = 2, \mathbf{Z}_{ik})$ is exponential with hazard $\gamma t^{\gamma-1} \exp(\beta_2' \mathbf{Z}_{ik})$, where $\beta_2 = \gamma \tau_2$. Each of $\beta_1, \beta_2, \tau_1, \tau_2$ is a $p \times 1$ vector.

Two designs are considered: cluster constant covariates design and matched design. In all cases, the model involves a scalar covariate, that is, $p = 1$. For each setup, data were generated repeatedly 5000 times, with the following algorithm:

- (i) Cluster sizes m_i are prespecified or randomly generated;
- (ii) For cluster constant design, covariates Z_{1i1} are randomly generated from a standard normal distribution for $i = 1, \dots, n$, with $Z_{1ik} = Z_{1i1}$ for all k ; for matched design, we let $Z_{1ik} = 0$ and $Z_{1ik'} = 1$, where k and k' are odd and even numbers, respectively;
- (iii) v_i are randomly generated from a positive stable distribution (α), where $\alpha \in (0, 1)$;
- (iv) h_i are randomly generated from a positive stable distribution (γ), where $\gamma \in (0, 1)$;
- (v) $2 - \varepsilon_{ik}$ are generated given v_i from a binary distribution with probability equaling $F_1(\infty | \mathbf{Z}_{1ik}, v_i) = 1 - \exp(-v_i e^{\tau_1' \mathbf{Z}_{1ik}})$;
- (vi) T_{ik} are generated from the conditional distribution of T_{ik} given ε_{ik} using inverse probability transformation.

4.2 Simulation results

We present simulation results for both cluster constant covariates design and matched design to demonstrate the performance of the clustered weighted (CW) score function relative to the clustered censoring complete (CCC) estimators and the estimators from the Fine–Gray model (described in section 3 of Fine and Gray, 1999).

For the cluster constant design, we conducted 2 batches of simulations to examine the sizes of Wald tests based on the different estimators of β_1 . In the first batch, cluster sizes are randomly generated with replacement from $\{2, 3, 4, 5\}$; in the second batch, the cluster sizes equaled 20 for all clusters. For each batch, we first generated data by assuming the true parameter values (β_1, β_2) to be $(0, 1)$ with independent standard normal covariates. The parameters for within-cluster correlations (α, γ) were assumed to be $(0.3, 0.3)$, $(0.3, 0.7)$, $(0.7, 0.3)$, or $(0.7, 0.7)$, and $\rho = 1$. Censoring times were independently generated from a uniform $[0.3, 1.5]$ distribution to achieve 20–30% of censored observations at various levels of (α, γ) .

Table 1 gives the empirical sizes of CC, CCC, weighted score (W) tests, and CW score tests for $n = 100$ and 250 and nominal level 0.05. Under all settings, the empirical sizes of tests not accounting for clustering (CC and W) deviate substantially from the nominal level. The CW tests and the CCC tests attain the nominal level when the number of clusters is reasonably large (250). When the number of clusters is smaller (100), their sizes are slightly larger than the nominal level.

Next, β_1 was set to 0.5 keeping all other parameters unchanged to assess the Wald tests under the alternative. The number of clusters is 100, with additional simulations performed with 250 and 500 clusters when $(\alpha, \gamma) = (0.3, 0.3)$ to assess the impact of sample size, for which the results using other settings are similar and are omitted.

Table 2 gives $E(\hat{\beta}_1)$, estimated with the average of the $\hat{\beta}_1$ from the 1000 replicates; $s(\hat{\beta}_1)$, estimated with the square root of the empirical variance of $\hat{\beta}_1$; and $E(\hat{s})$, the average of the model-based standard errors (SEs) of $\hat{\beta}_1$. The 2 clustered approaches (CCC, CW) have better performance than the unclustered approaches (CC, W), in terms of the empirical coverage of the 95% confidence intervals, and the closeness of model-based variance and empirical variances, under all settings. Note that when the within-cluster correlation is larger ($\alpha = 0.3$), the parameter estimate has slightly larger bias and the confidence intervals using the robust variance estimator have slightly reduced coverage versus when the correlation is smaller ($\alpha = 0.7$), all other things being the same. Larger numbers of clusters, 250 and 500, lead to estimators with smaller bias and corresponding empirical coverages which achieve the nominal level.

Table 1. Empirical sizes of tests from CC, CCC, W, and CW estimating equations with cluster-constant standard normal covariates. Cluster sizes $m \in \{2, 3, 4, 5\}$ or 20. $F_1(t; \mathbf{Z}_{ik}) = 1 - \exp\{-(1 - e^{-t})^\alpha e^{\beta_1' \mathbf{Z}_{ik}}\}$. Significance level = 0.05

α	γ	Censoring	Number of clusters	m	CC	CCC	W	CW
0.3	0.3	0.21	100	≤ 5	0.284	0.050	0.286	0.050
			250		0.275	0.056	0.276	0.056
0.3	0.7	0.22	100		0.284	0.050	0.286	0.050
			250		0.275	0.056	0.275	0.056
0.7	0.3	0.29	100		0.162	0.058	0.164	0.060
			250		0.176	0.056	0.171	0.055
0.7	0.7	0.30	100		0.162	0.058	0.165	0.060
			250		0.176	0.056	0.171	0.056
0.3	0.3	0.21	100	20	0.550	0.065	0.556	0.065
			250		0.549	0.056	0.552	0.055
0.3	0.7	0.22	100		0.550	0.065	0.557	0.065
			250		0.549	0.056	0.553	0.056
0.7	0.3	0.29	100		0.396	0.065	0.398	0.065
			250		0.396	0.053	0.397	0.052
0.7	0.7	0.30	100		0.413	0.058	0.416	0.057
			250		0.404	0.052	0.403	0.051

Table 2. Parameter estimates from CC, CCC, W, and CW estimating equations with cluster-constant standard normal covariates. Cluster sizes $m \in \{2, 3, 4, 5\}$. $F_1(t; \mathbf{Z}_{ik}) = 1 - \exp\{-(1 - e^{-t})^\alpha e^{\beta_1' \mathbf{Z}_{ik}}\}$

α	γ	Censoring	Number of clusters	β_1	Equation	$\hat{\beta}_1$	$s(\hat{\beta}_1)$	$E(\hat{s}_1)$	Coverage
0.3	0.3	0.22	100	0.5	CC	0.514	0.140	0.076	0.717
					CCC	0.514	0.140	0.134	0.938
					W	0.514	0.140	0.076	0.713
					CW	0.514	0.140	0.134	0.938
			250	0.5	CC	0.505	0.086	0.047	0.727
					CCC	0.505	0.086	0.085	0.944
					W	0.505	0.086	0.047	0.722
					CW	0.505	0.086	0.085	0.944
			500	0.5	CC	0.502	0.061	0.033	0.719
					CCC	0.502	0.061	0.060	0.948
					W	0.502	0.061	0.033	0.717
					CW	0.502	0.061	0.060	0.948
0.3	0.7	0.25	100	0.5	CC	0.514	0.140	0.076	0.717
					CCC	0.514	0.140	0.134	0.938
					W	0.514	0.140	0.076	0.712
					CW	0.514	0.140	0.134	0.938
0.7	0.3	0.30	100	0.5	CC	0.508	0.116	0.081	0.834
					CCC	0.508	0.116	0.112	0.938
					W	0.508	0.116	0.080	0.829
					CW	0.508	0.116	0.112	0.939
0.7	0.7	0.32	100	0.5	CC	0.508	0.116	0.081	0.834
					CCC	0.508	0.116	0.112	0.938
					W	0.508	0.116	0.080	0.829
					CW	0.508	0.116	0.112	0.939

For the matched design, cluster sizes were 2 and the true parameter values for β_1 were 0, 0.12, 0.24, 0.36, 0.48, or 0.60, and assuming β_2 to be 1. We took $(\alpha, \gamma, \rho) = (0.3, 0.6, 0.5)$ and $(0.6, 0.6, 2)$. Censoring times were independently generated from a uniform $[a, b]$ distribution with a and b giving censoring of 0%, 30%, and 50%.

Figure 1 depicts the powers of CC, CCC, and CW score tests at the nominal level of 0.05. Under all censoring percentages and both correlation levels, the unclustered tests have lower power than the clustered approach since the unclustered tests overestimate the variance of the parameter estimate (the corresponding model-based variance estimates are much bigger than the empirical estimates). The figure also suggests that the powers of CW tests and the CCC tests are almost the same, which is an indication of the validity of our weighting technique in the clustered case. Simulations with larger cluster sizes ($m_i = 10$, i.e. 5 in each arm) are also performed under the above settings, with similar results obtained.

It is worth highlighting that the simulation results are insensitive to the clustering parameter for the competing event (γ) under all settings.

5. APPLICATION

The Acute Myeloid Leukemia (AML) data arise from an ongoing bone marrow transplant registry of the European Blood and Marrow Transplant (EBMT) Group. In this analysis, the event of interest was the

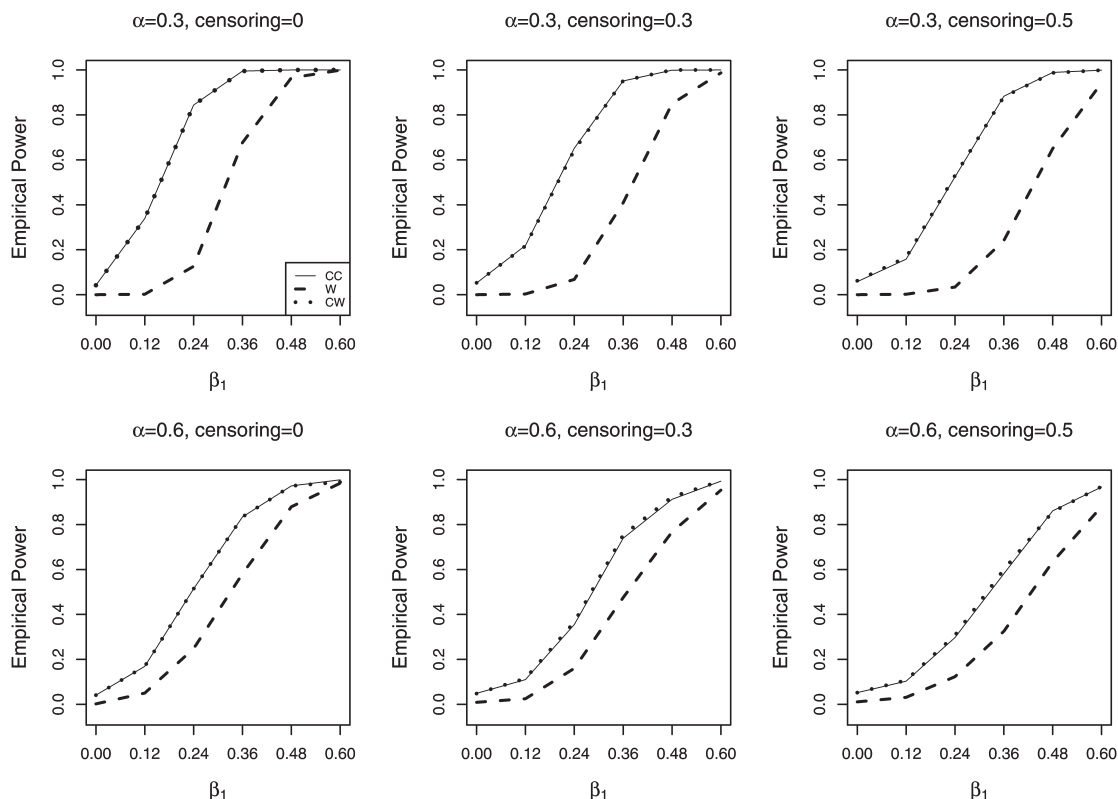


Fig. 1. The empirical powers of the tests for matched design.

time from graft to the first occurrence of either acute GvHD grade 2 or chronic GvHD. Death and relapse without GvHD are the competing causes of failure. Katsahian *and others* (2006) proposed a frailty model for the subdistribution hazard in order to test the prognostic factors while treating the centers as clusters. A subset of the data was used with the extraction date being January 1, 2002 consisting of patients with the following inclusion criteria: (1) received either genoidentical or matched unrelated donor (MUD) stem cell transplant; (2) were more than 16 years old at the time of transplant; (3) had acute myeloid leukemia in first complete remission; (4) received a transplant between January 1, 1994 and December 31, 2004; and (5) did not receive a reduced intensity regimen nor a T-cell-depleted transplant. Centers with only one patient enrolled were excluded. A total of 1022 patients from 121 clusters were included in their analysis using a frailty model.

In our analysis, we used the same registry, but with data extracted up to July 2008, while keeping other inclusion criteria from Katsahian *and others* (2006). The median follow-up was 1250 days, comprising patients still alive without relapse and disease. We have a total of 2952 patients from 244 centers, with 1385 GvHD and 629 competing causes of failure observed. The median number of patients per center was 6.

Since the patient populations are remarkably different across centers (Katsahian *and others*, 2006), there might be unobserved factors that are shared by patients within centers, potentially invalidating the use of inferences, which assume independence within centers. To address such clustering, model (2.1) is considered, where the covariates are the same as those in Katsahian *and others* (2006). These covariates, which are major predictive prognostic factors of GvHD, are defined as follows:

$\mathbf{Z}_{ik} = (Z_{1ik}, Z_{2ik}, Z_{3ik}, Z_{4ik})$ for the k th subject in the i th center, where $i = 1, \dots, 244, k = 1, \dots, m_i$, and $m_i \in \{2, \dots, 92\}$. Here, $Z_{1ik} = I(\text{female donor to male recipient [FM]})$, $Z_{2ik} = I(\text{source of stem cells is peripheral blood})$, $Z_{3ik} = I(\text{French-American-British [FAB] classification of AML is M5, M6, or M7})$, and $Z_{4ik} = I(\text{type of transplant is MUD})$. $I(A) = 1$ if A is true and 0 otherwise. We first fit univariable models (2.1), followed by a multivariable analysis.

The results of the coefficient estimates along with the SEs from the clustered model (robust) and the unclustered model (naïve) are reported in Table 3. In the univariable and multivariable analyses, both the Fine–Gray model and the proposed marginal model indicate gender matching between donor and recipient (female donor to male recipient vs. others) is a significant prognostic factor in the subdistribution hazard of GvHD occurrence at 0.01 significance level. FAB classification of AML is significant at 0.1 level under clustered multivariable analysis but insignificant under all 3 other models. Despite the similar results for the clustered and unclustered approaches, the naïve and robust variance estimators are quite different for some factors, suggesting that patients are correlated within cluster.

The estimated cumulative incidence function for 4 hypothetical patients are plotted in Figure 2. Here, we consider the source of the stem cells to be the peripheral blood, and that the transplant is from an unrelated donor. We further employed all possible combinations of the 2 significant covariates: FM and FAB classification of AML being M5, M6, or M7 (FAB1). One sees clearly that FM has a greater impact on the cumulative incidence of GvHD than does FAB classification, as evidenced by its impact on the absolute probabilities of GvHD.

Table 3. *Parameter estimates in models for acute GvHD or chronic GvHD occurrence*

	Univariate			Multivariable		
	$\hat{\beta}$	SE (naïve)	SE (robust)	$\hat{\beta}$	SE (naïve)	SE (robust)
Female to male versus others	0.363	0.060 [†]	0.065 [†]	0.368	0.063 [†]	0.068 [†]
PBSC versus BMT	−0.051	0.054	0.075	−0.060	0.057	0.076
FAB M5, M6, M7 versus others	0.089	0.067	0.058	0.099	0.069	0.060 [‡]
MUD versus genoidential	0.049	0.081	0.103	0.086	0.087	0.108

[†]Significant at 0.01 level.

[‡]Significant at 0.1 level.

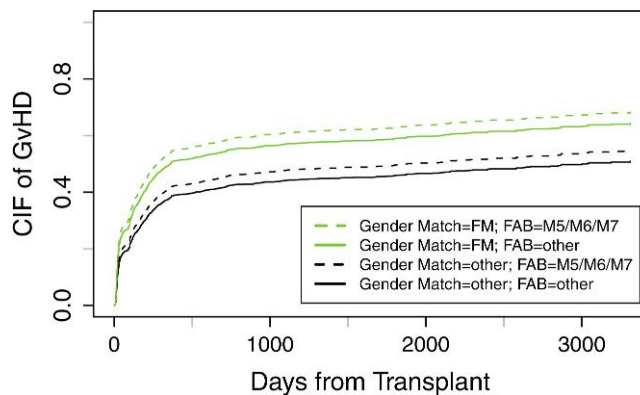


Fig. 2. The estimated cumulative incidence function.

6. DISCUSSION

The use of marginal models has been widely adopted for ordinary right-censored data. The proposed methods provide an adaptation of the Fine–Gray model for the cumulative incidence function, which rigorously accommodates both correlated failure times and correlated censoring times. Such methods are particularly useful in applications with small groups of correlated observations where the correlation is primarily a nuisance, as in multicenter trials. Frailty models are less attractive in such settings, owing to the need to explicitly model such correlations, which complicates the analysis of covariate effects.

The simulations demonstrate the potential bias and loss of power in hypothesis testing, which may arise from ignoring within-cluster correlations in variance estimation. Additional improvements in power might be achieved via more careful consideration of the correlation structure. For example, model-based approaches, like frailty models (Katsahian *and others*, 2006), might potentially yield such gains, at the risk of bias under model misspecification. The development of tests which yield increased power while still being robust to misspecification is a topic of future research.

An R function, CRRC, which implements the marginal analysis, is included in the supplementary material available at *Biostatistics* online and has been incorporated in crrSC, an R package which is publicly available on the Comprehensive R Archive Network (CRAN) site.

SUPPLEMENTARY MATERIAL

Supplementary material is available at <http://biostatistics.oxfordjournals.org>.

ACKNOWLEDGMENTS

The authors thank Vanderson Rocha at EBMT Group for permission to use the EBMT registry data.
Conflict of Interest: None declared.

FUNDING

National Institutes of Health (NIH)/National Cancer Institute (NCI) (1R01 CA94893-02 to J.F.).

REFERENCES

- CAI, J. W. AND PRENTICE, R. L. (1997). Regression estimation using multivariate failure time data and a common baseline hazard function model. *Lifetime Data Analysis* **3**, 197–213.
- CHEN, B. E., KRAMER, J. L., GREENE, M. H. AND ROSENBERG, P. S. (2008). Competing risks analysis of correlated failure time data. *Biometrics* **64**, 172–179.
- CLAYTON, D. AND CUZICK, J. (1985). Multivariate generalizations of the proportional hazards model. *Journal of the Royal Statistical Society. Series A (General)* **148**, 82–117.
- FINE, J. P. AND GRAY, R. J. (1999). A proportional hazards model for the subdistribution of a competing risk. *Journal of the American Statistical Association* **94**, 496–509.
- GRAY, R. J. (1988). A class of K-sample tests for comparing the cumulative incidence of a competing risk. *Annals of Statistics* **16**, 1141–1154.
- HOUGAARD, P. (1986). Survival models for heterogeneous populations derived from stable distributions. *Biometrika* **73**, 387–396.
- KATSAHIAN, S., RESCHE-RIGON, M., CHEVRET, S. AND PORCHER, R. (2006). Analysing multicenter competing risks data with a mixed proportional hazards model for the subdistribution. *Statistics in Medicine* **25**, 4267–4278.

- LEE, E. W., WEI, L. J. AND AMATO, D. A. (1992). Cox-type regression analysis for large numbers of small groups of correlated failure time observations. In: Klein, I. P. and Goel, P. K. (editors), *Survival Analysis: State of the Art*. Dordrecht: Kluwer Academic Publishers. pp. 237–247.
- LIANG, K. Y., SELF, S. G. AND CHANG, Y. C. (1993). Modelling marginal hazards in multivariate failure time data. *Journal of the Royal Statistical Society. Series B, Statistical Methodology* **55**, 441–453.
- LOGAN, B. R., ZHANG, M. J. AND KLEIN, J. P. (2011). Marginal models for clustered time to event data with competing risks using pseudovalues. *Biometrics* **67**, 1–7.
- PRENTICE, R. L., KALBFLEISCH, J. D., PETERSON, JR, A. V., FLOURNOY, N., FAREWELL, V. T. AND BRESLOW, N. E. (1978). The analysis of failure times in the presence of competing risks. *Biometrics* **34**, 541–554.
- ROBINS, J. M. AND ROTNITZKY, A. (1992). Recovery of information and adjustment for dependent censoring using surrogate markers. In: Jewell, N., Dietz, K., and Farewell, V. (editors), *AIDS Epidemiology—Methodological Issues*. Boston, MA: Birkhäuser. pp. 297–331 (includes errata sheet).
- RUAN, P. AND GRAY, R. J. (2008). Analyses of cumulative incidence via parametric multiple imputation. *Statistics in Medicine* **27**, 5709–5724.
- SPIEKERMAN, C. F. AND LIN, D. Y. (1998). Marginal regression models for multivariate failure time data. *Journal of the American Statistical Association* **93**, 1164–1199.
- WEI, L. J., LIN, D. Y. AND WEISSFELD, L. (1989). Regression analysis of multivariate incomplete failure time data by modeling marginal distributions. *Journal of the American Statistical Association* **84**, 1065–1073.
- YING, Z. AND WEI, L. J. (1994). The Kaplan–Meier estimate for dependent failure time observations. *Journal of Multivariate Analysis* **50**, 17–29.

[Received October 4, 2010; revised September 1, 2011; accepted for publication September 2, 2011]